AD-A080 559 UNIVERSITY OF CENTRAL FLORIDA ORLANDO DEPT OF MATHEM--ETC F/G 4/2

STOCHASTIC MODELING OF CLIMATIC PROBABILITIES.(U)

NOV 79 P N SOMERVILLE, S J BEAN F19628-77-C-0080

UNCLASSIFIED AFGL-TR-79-0222 NL

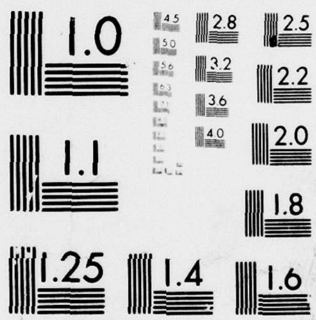
1 OF 1

AD
A080559

END

DATE
FILMED
3-80

DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12) LEVEL II

AFGL-TR-79-0222

Stochastic Modeling of Climatic Probabilities

Paul N. Somerville

Steven J. Bean

Department of Mathematics and St.tistics

University of Central Florida

Orlando, FL 32816

Final Report

1 June 1977 - 31 August 1979

November 1979

Approved for Public Release; Distribution Unlimited

Air Force Geophysics Laboratory

Air Force Systems Command

United States Air Force

Hanscom AFB, Massachusetts 01731

Qualified requestors may obtain additional copies from the Defense Documentation Center. All others should apply to the National Technical Information Service.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(19) REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (18) AFGL-TR-79-0222

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER (9)

4. TITLE (and Subtitle) (6) STOCHASTIC MODELING OF CLIMATIC PROBABILITIES.

5. TYPE OF REPORT & PERIOD COVERED Final Rep[redacted]t. 1 Jun[redacted]77-31 Aug[redacted]79

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s) (10) Paul N. Somerville, Steven J. Bean

8. CONTRACT OR GRANT NUMBER(s) (15) F19628-77-C-0080

9. PERFORMING ORGANIZATION NAME AND ADDRESS
University of Central Florida
Department of Mathematics and Statistics
P.O. Box 25000, Orlando, FL 32816
411 423

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) 62101 F, 667009 AD (17) 09

11. CONTROLLING OFFICE NAME AND ADDRESS
Air Force Geophysics Laboratory
Hanscom AFB, Massachusetts 01731
Monitor/Irving I. Gringorten/LYD

12. REPORT DATE (11) Nov[redacted]79

13. NUMBER OF PAGES

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 22

15. SECURITY CLASS. (of this report) Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release, Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Modeling	Rainfall	S-Curve
Distributions	Ceiling	Extreme Value
Visibility	Climatic Probabilities	Kappa
Windspeed	Burr Curve	
Skycover	Weibull	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report summarizes the work done in the past two years in the Stochastic Modeling of Climatic Probabilities of weather events. Models are presented for visibility, windspeed, skycover, rainfall and ceiling. The models were each applied to more than twenty stations.

SECURITY CLASSIFICATION OF THIS PAGE(*When Data Entered*)

SECURITY CLASSIFICATION OF THIS PAGE(*When Data Entered*)

CONTRIBUTORS

Dr. Paul N. Somerville was the Principal Investigator. Dr. Steven Bean was a major contributor during the second year of the investigation. Harold Griffith was available as a consultant during the entire period and his suggestions and advice were most helpful.

Students who made major scientific or technical contributions to the project and whose names appear as co-authors of one or more technical reports are Richard Daley, Sherill Falls, Mark Heuser and Sharon Watkins. Other students who contributed in a major way to the success of the project are Sarah Autrey, Jeff Einerson, Karl Grammel, Tom Ticknor and Debbie Waite.

A special expression of appreciation is necessary for Ann Henry, Lois Voge and Patti Martin for their careful and painstaking efforts in typing and proofreading of the scientific reports and the final report.

TABLE OF CONTENTS

0. INTRODUCTION

Within the last few years there has been an increasing public awareness of the impact of weather and climate on mankind. The Air Weather Service of the U.S. Air Force has had a continuing and exponentially growing mission to provide climatological information to weapon system designers and operators, and to planners on both the strategic and tactical level. The increasing sophistication and cost of weapons systems and the recognition that the environment degrades or offers opportunities has led to the requirement for more and better climatological information.

The present study was motivated by the goal of achieving a capability to determine the climatic probability of above-threshold conditions of weather relative to the success of an Air Force flight mission, anywhere, at any time expeditiously. A more limited goal of this research was to "expedite the determination of the climatic probability of any desired weather event, anywhere in the world, for any hour of the day."

The study focused on five weather elements - visibility, windspeed, skycover, rainfall and ceiling. For each of these, useful probabilistic models were obtained, and applied to more than twenty stations.

1. METHODOLOGY

Vast amounts of data are available from many worldwide locations and for many weather elements. A wealth of data is stored in Asheville, North Carolina, "the largest climatic center in the world." To make some of the data more accessible RUSSWO's (Revised Uniform Summary of Weather Observations) and SMOS's (Summary of Meteorological Observations, Surface) have been prepared by the Data Processing Division of the Air Weather Service and the Naval Weather Service Detachment respectively.

To estimate the probability of some weather event such as the probability of a windspeed less than 5 knots at Bedford at 1300 hours on July 4, 1984, one could look up the original records or summaries for 1300 hours on previous July dates, and estimate the required probability from these past records. To have the capability of obtaining predictions for many weather elements, for many locations, for arbitrary months and times of day obviously would require access to a voluminous amount of data.

An alternate method of estimation of the required probabilities is by the use of probability models. A very elementary method of developing a model for data is the following. First, make a histogram of the data, and then "smooth" the histogram to obtain a frequency distribution (probability density function). The probability of a value of the variable less than some stated amount is then estimated by the proportion of the area under the frequency distribution to the left of that amount.

There are usually a number of curves or distributions which can be used to "fit" the data. If there are theoretical justifications for specific distributions, then, of course, those should be used. Where there are no theoretical justifications, a number of families of empirical distributions can be used. Johnson, Pearson and Burr families of curves each are capable of describing a wide variety of frequency distributions. The Pearson curves are probably the oldest and best known. An advantage of Johnson curves is that estimates of the percentiles of the fitted distribution can be obtained using a table of areas under a standard normal distribution. Burr curves have a closed form cumulative distribution function. That is, the probability that a random variable has a value less than some specified amount can be calculated by substitution in a simple expression, with no need for tables or numerical integration.

Once a particular distribution or family of distributions has been selected, for a given set of data, the parameters must be estimated. For example, the frequency distribution function for the gamma distribution is given by

$$f(x) = a^b x^{b-1} e^{-ax} / \Gamma(b) \quad \text{for } x \geq 0,\ a,\ b > 0$$
$$= 0 \quad \text{elsewhere}$$

The parameters a and b must be chosen so that the frequency distribution curve best fits the data.

The usual (and efficient) method for estimation of the parameters is by the method of maximum likelihood, or in some cases by the somewhat less efficient method of moments. With the parameters thus estimated, probabilities of conditions above or below a certain level can be obtained by integrating the probability density function.

We have taken a slightly different position in our modeling. First, we believe that the probability density function itself is of limited interest. What we are usually interested in is the probability that the variable of interest will exceed, or be less than some stated value. That is, it is the cumulative distribution function (c.d.f.) that is of greatest interest. With this in mind it seemed logical to investigate the possibility of restricting our first choice of probability density functions to those whose c.d.f.'s are in closed form. In this way we avoided the need for numerical integration or tables.

The idea of fitting distributions, which have closed form cumulative distribution functions, is not new. I.W. Burr (1942) proposed their use. However, even his "Type XII" distribution (Burr Curve) has only recently received any notice in the literature.

We have been able to find closed form cumulative distribution functions which give good models for the weather elements of interest.

There is a second area in which we have departed from the traditional modeling techniques. Instead of estimating the parameters of the probability density function using maximum likelihood estimators, or the method of moments we have chosen a different technique. We have regressed the empirical cumulative distribution on the model cumulative distribution. The model parameter estimates are thus those which minimize the sums of squares of the difference between the "corners" of the empirical c.d.f. and the corresponding points on the model cumulative distribution function.

We believe the above procedures are superior for our problem. We are interested in the probabilities that a weather element will be below (or above) a given level. The parameter estimates are those for

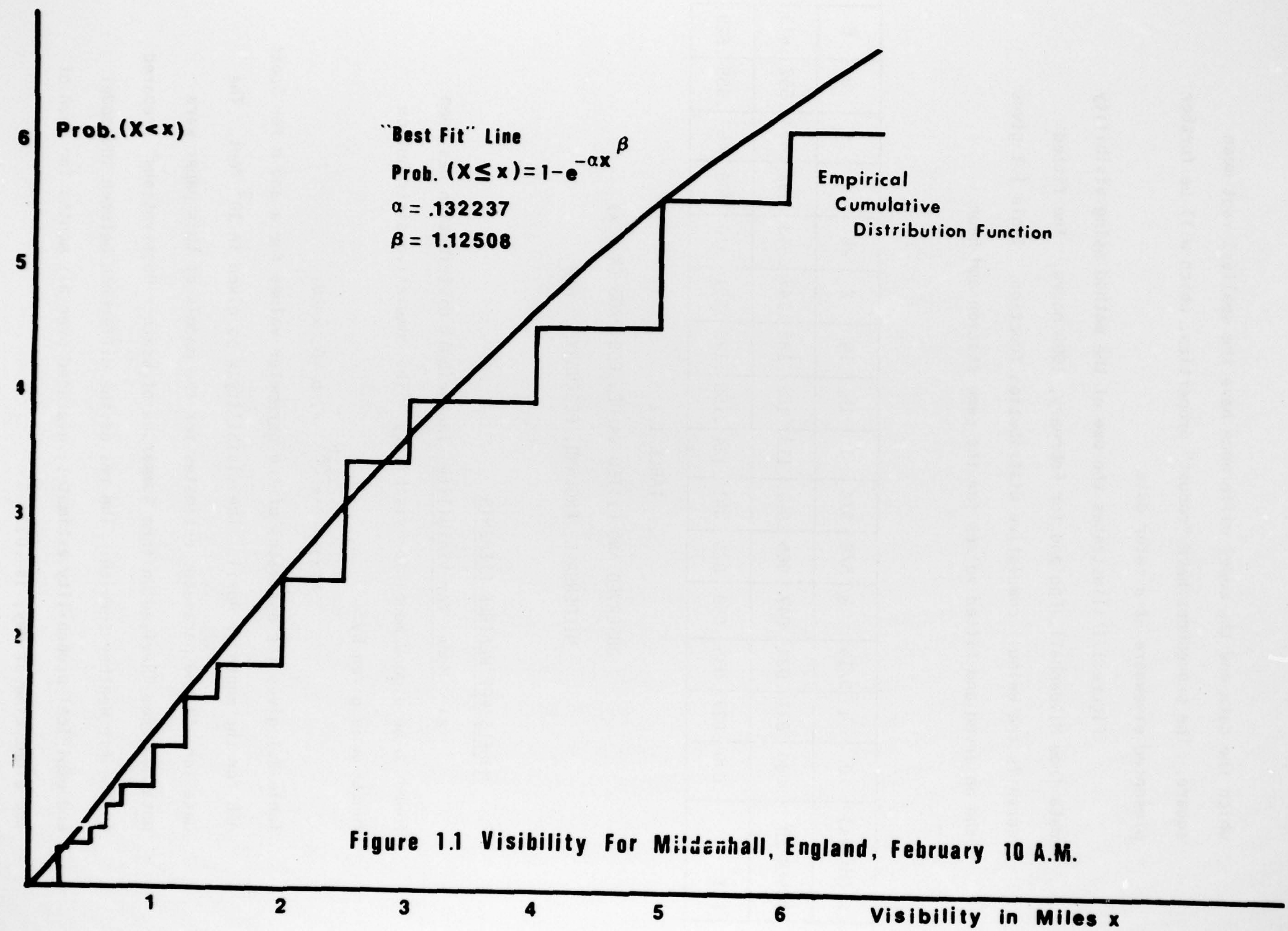


Figure 1.1 Visibility For Mildenhall, England, February 10 A.M.

which the data and the model difference have the smallest root mean square. The procedures have "robust" properties, which will be further presented elsewhere at a later date.

Figure 1.1 illustrates the use of the method using visibility data from Mildenhall, England for February, 1000 hours. The fitted curve is the Weibull cumulative distribution function. Table 1.1 gives the observed and fitted values for the same station and hour.

x (Miles)	0	¼	5/16	½	5/8	3/4	1	1¼	1½	2	2½	3	4	5	6
OBSERVED	.000	.031	.034	.047	.065	.081	.113	.152	.180	.247	.343	.392	.453	.557	.613
FIT	.000	.027	.035	.059	.075	.091	.124	.156	.188	.251	.310	.366	.467	.555	.629

TABLE 1.1

OBSERVED AND FITTED VALUES FOR PROB ($X \leqq x$)

MILDENHALL, ENGLAND, FEBRUARY 10 A.M.

2. MODELS FOR WEATHER ELEMENTS

a) Model for Visibility The Weibull distribution has been found to be a good model for visibility. The cumulative distribution function is given by

$$F(x) = 1-e^{-ax^b} \quad a, b > 0, x \geq 0$$

Table 2.1 gives the estimates of the parameter values for a and b for Scott AFB for the month of April. The visibility x is given in 10^3 feet. The data for all the parameter estimates for the models of this paper were obtained from "Revised Uniform Summaries of Weather Observations" prepared by the Air Weather Service. The rms of the difference between the model and empirical probability estimates, averaged over all months for each of the eight timesof day, is .01.

TIME OF DAY								
	0100	0400	0700	1000	1300	1600	1900	2200
a	.00217	.00536	.00934	.00217	.00106	.00152	.00202	.00163
b	2.504	2.169	1.973	2.406	2.602	2.420	2.393	2.528

TABLE 2.1

ESTIMATES OF VISIBILITY

TABLE OF SCOTT AFB PARAMETERS FOR WEIBULL MODEL

MONTH OF APRIL

A more detailed description of the methodology and tables of parameter values for twenty two diverse locations by months and time of day is given in Scientific Report #3, "Some Models for Visibility," AFGL-TR-79-0144 of this contract.

b) Model for Sky Cover A model commonly used for sky cover is the Beta distribution. We propose a new distribution, which we call the S distribution. The cumulative distribution function s given by

$$F(x) = 1 - (1-x^a)^b \quad a, b>0, 0\leqq x\leqq 1$$

The random variable x is the proportion of the sky covered by clouds. in a forthcoming paper the authors demonstrate that for nearly all Beta distributions, there is an S distribution function which is a very close approximation.

Table 2.2 gives the estimates of the parameter values for a and b for Balboa, Canal Zone for the month of May. The rms of the difference between the model and empirical probability estimates averaged over all months, for each of the eight times of day is .017.

	TIME OF DAY							
	0100	0400	0700	1000	1300	1600	1900	2200
a	.511	.654	1.740	2.601	3.206	3.574	1.375	.586
b	.385	.415	.353	.387	.340	.269	.226	.310

TABLE 2.2

ESTIMATES OF SKYCOVER PARAMETERS

S-DISTRIBUTION, BALBOA, MONTH OF MAY

A more detailed description of the methodology and tables of parameter values for twenty-three diverse locations by month and time of day is given in Scientific Report #2, "Some Models for Skycover," AFGL-TR-78-0219 and in Scientific Report #5, "A New Model for Skycover," AFGL-TR-79-0219. Both were written as a part of this contract.

c) Model for Windspeed The Weibull model has been found to be a good model for windspeed. Assuming the probability of "calm" to be c, then we use the cumulative distribution function

$$F(x) = c + (1 - c)(1 - e^{-ax^b}) \quad a, b, c>0, x\geqq 0$$

We thus use a three parameter model. Table 2.3 gives the estimates of the parameter values for a, b, and c for Hill AFB for the month of November. The windspeed x is given in knots. The rms of the difference between the model and empirical probability estimates averaged over all months, for each of the eight times of the day is .01.

	TIME OF DAY							
	0100	0400	0700	1000	1300	1600	1900	2200
a	.0112	.008[illegible]	.0086	.0102	.0233	.0268	.0283	.0166
b	2.049	2.10[illegible]	2.044	2.065	1.777	1.797	1.772	1.980
c	0.105	0.08[illegible]	0.083	0.101	0.162	0.193	0.251	0.163

TABLE 2.3

TABLE OF ESTIMATES OF WINDSPEED PARAMETERS

FOR WEIBULL DISTRIBUTION, HILL AFB

MONTH OF NOVEMBER

A more detailed description of the methodology and tables of parameter values for windspeed for twenty five diverse locations by month and time of day is given in Scientific Report #4, "Some Models for Windspeed", AFGL-TR-79-0180 of this contract.

d) <u>Model for Daily Precipitation</u> Mielke (1973) has used the two parameter kappa distribution for rainfall, and introduced the three-parameter kappa distribution. The cumulative frequency distribution for the three-parameter distribution is given by

$$F(x) = [(x/b)^{at}/(a+(x/b)^{at})]^{1/a} \quad a,b,t>0 \;\; x \geq 0.$$

The daily rainfall x is measured in inches. Although there is a positive probability of no precipitation, the kappa distribution can be used to obtain the estimated probability of no precipitation (including trace) if we put x = .005 inches.

Table 2.4 gives the estimates of the parameter values for a, b and t for Patrick Air Force Base for January, February, and March.

MONTH	a	b	t	RMS of Fit
Jan	40	.100	.05	.015
Feb	110	.975	.05	.011
Mar	100	.800	.05	.009

TABLE 2.4

TABLE OF ESTIMATES OF

KAPPA DISTRIBUTION PARAMETERS

FOR DAILY PRECIPITATION PATRICK AFB

JAN., FEB., MAR.

A more detailed description of the methodology and tables of parameters for daily precipitation by month are given in Scientific Report #1, "Some Models for Rainfall", AFGL-TR-78-0218 and in Scientific Report #6, "Some Additional Models for Rainfall", AFGL-TR-79-0220, both written as a part of this contract.

e) <u>Model for Ceiling</u> The Burr Curve was utilized for modeling ceiling. The cumulative distribution for the Burr Curve is given by

$$F(x) = 1 + (1 - (x/c)^a)^{-b} \quad a,b,c > 0,\ x \geqq 0.$$

Table 2.5 gives estimates of the parameter values a, b and c for Bangor, Maine, for the month of May. The rms of the difference between the model and the empirical probability estimates averaged over all hours, for May, is 0.2.

	TIME OF DAY							
	0100	0400	0700	1000	1300	1600	1900	2200
a	.6123	.6183	.9415	1.3016	1.9219	1.6175	.8940	.6723
b	.7378	.6199	.2674	.2510	.1930	.2442	.7729	.7039
c	10000	5000	1000	1500	1500	2000	10000	10000

TABLE 2.5

ESTIMATES OF "BURR CURVE" PARAMETERS FOR CEILING

BANGOR, MAINE, MAY

A more detailed description of the methodology and tables of parameter values for ceiling for twenty-three diverse locations by month is given in Scientific Report #7, "Some Models for Ceiling", AFGL-TR-79-0221.

f) Extreme Value Model for Rainfall The "extreme value" distribution was used for modeling the maximum amount of daily rainfall for a specified month. The cumulative distribution function is given by

$$F(x) = \exp(-\exp(-((x-\mu)/\sigma))) \qquad \sigma > 0$$

where μ and σ are constants or the parameters for the distribution.

Table 2.6 gives the estimates m and s of the parameter values μ and σ for Patrick AFB, and the rms values for each month.

	m	s	RMS
Jan.	0.553	0.833	0.034
Feb.	0.913	0.770	0.051
Mar.	0.780	0.757	0.030
Apr.	0.584	0.563	0.030
May	0.885	0.787	0.025
Jun.	1.239	0.964	0.024
Jul.	1.018	0.404	0.033
Aug.	1.276	0.562	0.039
Sep.	1.828	1.034	0.038
Oct.	1.201	1.155	0.033
Nov.	0.653	0.858	0.048
Dec.	0.524	0.479	0.019

TABLE 2.6

ESTIMATES OF "EXTREME VALUE" PARAMETERS

MAXIMUM DAILY RAINFALL

PATRICK AFB

A more detailed description of the methodology and tables of parameter for nine diverse locations are given in Scientific Report #1, "Some Models for Rainfall", AFGL-TR-78-0218, written as a part of this contract.

g) <u>Overall or Comprehensive Models</u> For skycover and daily precipitation, some "overall" models were developed. By this we mean models in which the month of the year m (1, 2, ..., 12) and/or the hour of the day h (1, 2, ..., 24) are included in the formula giving the probability (cumulative distribution function) of a value of skycover (or rainfall) less than some stated amount.

As an example for Patrick AFB, we may use the following formulas for a, b and t.

$$a = 61.59 + 7.880\,m - .652\,m^2$$

$$b = .0095 + .361\,m - .029\,m^2$$

$$t = .021 + .015\,m - .011\,m^2$$

where m is the month of the year. The formulas in effect replace the table of parameter values for the station (e.g. Table 2.4) The rms of the error (over all months) using the above formulas is .01, compared to the values .015, .011, .009 for January, February and March using separate parameter values for each month.

Scientific Report #1, "Some Models for Rainfall", contains "overall" rainfall models for the stations modeled there. The parameter estimates for some models are polynomials in m and in others include sine or cosine functions.

Scientific Report #2 "Some Models for Skycover" contains "overall skycover models for the stations modeled there. Since for skycover we have different models for months and times of day, the "overall" models have parameter estimates which are functions of both month m, and time of day h. Models given in the report include polynomials in m and h, and models including sine and cosine terms in addition to the polynomial terms.

"Overall" or "comprehensive" models were not developed for other weather elements. The advantage of the "overall" model is that the requirement for a table of values for each month and/or time of day no longer exists. The table is replaced by a set of formulas. Overall models have some disadvantages. First, the overall models are less accurate. They can be made more accurate by increasing the number of terms in the formula, but that in turn increases the amount of calculation, and also the amount of storage required in a computer program, and may cancel out the advantage of not needing a table.

Although, we doubt that in general, "overall" models will be preferable to "individual" models, we have demonstrated that they can be developed, and have developed some "overall" models for skycover and rainfall.

3. SUMMARY OF ACCOMPLISHMENTS

It has been demonstrated that models can be developed to compact historical records of weather events. Models have been developed for visibility, skycover, windspeed, rainfall and ceiling. These models have been used to "compact" historical records for each of the above weather elements for more than twenty stations at diverse locations. The following models (distributions) have been successfully utilized.

Visibility	-	Weibull
Skycover	-	Beta, Johnson Curve, S-curve
Windspeed	-	Weibull
Rainfall	-	Lognormal, Three-parameter Kappa
Ceiling	-	Burr Curve

It has been demonstrated that "overall" or "comprehensive" models can be developed which eliminate the need for separate tables of parameter values for different months and/or time of day. These models are not as accurate, but accuracy can be increased by adding more terms for the formulas for the parameter estimates. This is at the expense of additional complexity, computation and storage. "Overall" models for skycover and rainfall were developed and applied to a limited number of stations.

Departing from some standard methodologies, the principle of modeling using distributions having closed form cumulative distribution functions was adopted. Since the objective was the estimation of probabilities, this means that the required probabilities can in all cases be calculated from straightforward formulas with no need for additional approximations or numerical integration.

Another departure from traditional modeling techniques was the estimation of the model parameters by regressing the empirical cumulative distribution function on the model distribution function.* The model parameter estimates are thus those which minimize the sums of squares of the "differences" between the empirical and model distribution functions (probabilities). The optimum properties of the procedure is the subject of a planned future publication.

A new distribution, referred to as the S-distribution was introduced, and it was used for modeling skycover. A discussion of the properties of the S-distribution is the subject of a planned future publication.

The following Scientific Reports were published.

#1	Some Models for Skycover	TR-1978-0218
#2	Some Models for Rainfall	TR-1978-0219
#3	Some Models for Visibility	TR 1979-0144
#4	Some Models for Windspeed	TR-1979-0180
#5	A New Model for Skycover	TR-1979-0219
#6	Some Additional Models for Rainfall	TR-1979-0220
#7	Some Models for Ceiling	TR-1979-0221

*A more detailed description of the method is planned for a future Scientific Report entitled, "Use of Non-linear Regression to Estimate a Cumulative Distribution Function."

4. REFERENCES

Burr, I.W., 1942: "Cumulative frequency functions"
Annals of Math. Stat., 13, pp. 233-235.

Hahn, G.J. and S.S. Shapiro, 1967: Statistical
Models in Engineering
John Wiley and Sons, N.Y.

Mielke, Paul W. Jr., 1973: "Another family
of distributions for describing and
analysing precipitation data"
J. Appl. Meteor., 12 pp. 275-280.